Interfacing Lexical and Ontological Information in a Multilingual Soccer FrameNet

Thomas Schmidt

International Computer Science Institute
1947 Center Street, Berkeley, CA 94704
Schmidt@icsi.berkeley.edu

Abstract

This paper presents ongoing work on a multilingual (English, French, German) lexical resource of soccer language. The first part describes how lexicographic descriptions based on frame-semantic principles are derived from a partially aligned multilingual corpus of soccer match reports. The remainder of the paper then discusses how different types of ontological knowledge are linked to this resource in order to provide an access structure to the resulting dictionary. It is argued that linking lexical resources and ontologies in such a way provides novel ways to a dictionary user of navigating a domain vocabulary.

1. Introduction

This paper presents ongoing work on a multilingual lexical resource of soccer language – the Soccer FrameNet (SFN). At present, three languages – French, (British) English and German – are taken into account, but the design is potentially open to include additional languages (see also section 1.2). The overall goal is to organize verbs, nouns, adjectives and idiomatic expressions that are used to describe actors, objects and events in and around a soccer match into a lexical network. This network should then serve as an electronic (mono- or bilingual) dictionary to a human user and – potentially – be exploitable by a machine for purposes of natural language processing (e.g. semantic web technology).

The methodological starting point for the development of the lexical resource is frame semantics (Fillmore 1982) and the methodology employed in the construction of the FrameNet lexicon (Ruppenhofer et al. 2005). This basically means that the notion of the semantic frame – "a script-like conceptual structure that describes a particular type of situation, object or event and the participants and props involved in it" (Ruppenhofer et al. 2005) – is used as a fundamental organization principle of the lexicon above the individual linguistic unit. As has been argued, for instance in (Boas 2005), semantic frames can also act as a kind of interlingua for multilingual resources.

The paper is structured as follows: Section 3 gives an overview of the project and illustrates how the basic lexical descriptions of the resource are organised. The following sections then discuss ways of providing additional structure to these lexical descriptions on the basis of ontological-driven principles. Section 5 discusses the assignment of lexical units to a poly-hierarchy of concepts, section 6 shows how arguments of lexical units can be linked to an ontology in the same way, and section 7 introduces the notion of a scenario as an additional ontological structure that can help to organize the lexicon.

2. Related Work

Soccer has been chosen as an exemplary domain in a number of studies related to ontologies as well as in lexicographic research.

Regarding the latter, several contrastive (mostly French-German) analyses of soccer vocabulary have been carried out in the framework of lexicon grammar, most notably by Seelbach (2001, 2002 and 2003). The project presented here differs from that work not only in the choice of languages (English, German and French) and in the basic theoretical approach (frame semantics), but also in its effort to go beyond an exemplary analysis of a small number of examples and instead provide a comprehensive electronic lexical resource which covers a substantial part of the entire soccer vocabulary.

Regarding ontologies, the MUMIS project has constructed a soccer ontology for the purpose of multi-media retrieval of soccer data (Nijholt et al. 2003, Reidsma et al. 2003). Currently, the SMARTWEB project is developing a sports event ontology as a component of a cross lingual, cross media semantic web application for the soccer world cup 2006 in Germany (Buitelaar et al. 2005, Buitelaar et al. 2006). In these projects, the focus is clearly on machine processing of natural language, and the ontologies of these systems consequently play a much more central role than in this project, where the focus is on lexicographic description and ontologies are simply seen as *one* means of organizing such descriptions (see below).

3. Project overview

3.1. Design principles

Although, in constructing the SFN, frame semantics provides the basic methodology for the analysis and the representation of lexical descriptions, there are two reasons not to follow the guidelines for the development of the General Language FrameNet (GLFN) by the book: firstly, the GLFN methodology has been developed with a monolingual lexicon in mind and some requirements that arise only in the construction of a multilingual resource may consequently not have been taken into account.

Secondly, in contrast to the GLFN, the SFN is a domain specific resource. This restriction holds the potential for some methodological alterations. Most importantly, this regards the fact that the number of relevant lexical units will be limited to a comparatively low, finite number (not greater than 1,500 for each language, as a first careful estimate, see also section 3.4) making it possible for the lexicographer to maintain a much more complete and detailed overview of the resource than would be feasible in the general language case. A bottom-up approach to the organization of the lexicon – starting with a "flat" list of LUs and then adding structure to this list – as described in more detail below is greatly facilitated by this fact.

3.2. Some general characteristics of soccer language

All lexical units investigated so far fall into one of the following categories:

- soccer terms: words specifically coined for concepts in soccer, e.g. the noun 'free-kick' or the verb 'to wrong-foot' in English, the noun 'Strafstoß' or the verb 'dribbeln' in German, the noun 'coup de pied arrêté' or the verb 'tacler' in French;
- soccer jargon: words used also in general language, but taking on a distinctively specified meaning when used for talking about soccer, e.g. the noun 'wall' or the verb 'to save' in English, the noun 'Fahrkarte' or the verb 'tunneln' in German, the noun 'petit pont' or the verb 'expulser' in French;
- general language: words frequently used in soccer reports and not having a distinctively different meaning when used outside soccer, e.g. the noun 'victory' or the verb 'to lose' in English.

An obvious characteristic of soccer language, and one that makes it especially interesting for lexicographic purposes, is that it abounds with synonyms. More often than not, one and the same concept can be expressed by more than one lexical item. Consider for instance, the following collection of German verbs each of which can be used to describe that a player overcomes his opponent in a one-on-one challenge:

(1) ausdribbeln, ausspielen, austanzen, austricksen, düpieren, tunneln, umdribbeln, umspielen, verladen, vernaschen, versetzen

Likewise, it is very common in soccer reports to alternate between synonymous nominal and verbal predicates:

(2) Substitute Nilmar **was fouled** by Frank Fahrenhorst just inside the area.[1]
(3) Frank Fahrenhorst **commited a foul** on substitute Nilmar just inside the area.

1 All examples are authentic corpus examples but have been shortened for the purpose of this paper.

3.3. Corpus data

A partially aligned corpus of soccer match reports is used to carry out the lexicographic analysis. The core corpus consists of approximately 500 texts (coming up to around 300,000 words) in each of the languages English, German and French. Around half of these texts are parallel – i.e. they are direct translations of one another –, while the other half consists of comparable texts – i.e. they report the same match but have been written independently of one another. All of the texts have been retrieved from the official website of the UEFA (www.uefa.com). This core corpus is supplemented by additional material from other sources. For German, this comprises match reports from a German soccer journal (www.kicker.de) amounting to roughly 1,000,000 words. For English and French, there are altogether 200,000 more words from other sources. The UEFA website also contains soccer reports in Spanish, Italian, Portuguese, Russian and Japanese. These have also been retrieved in the acquisition process and could potentially be used to supplement the resource for other languages in the future.
All texts have been preprocessed: for the core corpus, this involved tokenizing and sentencizing the text, identifying hyphenated compounds and other automatically detectable multi-word expressions, as well as aligning the parallel portions of the corpus on the paragraph level. All texts are stored in TEI compliant XML.

3.4. Lexicographic data

On the most basic level, the development of the lexical resource consists in finding usages of soccer specific lexical units (like "header", "offside", "to nutmeg", "to defeat") in the corpus, to analyze their argument structure following frame semantic principles, to write a definition that incorporates this argument structure analysis and to annotate a number of example sentences for each unit according to this analysis. The following are examples of resulting LU descriptions for the English noun "cross" the English verb "to dispossess" and the German verb "tunneln":

cross.n

Using a part of his body (ARG4), a player (ARG1) transfers the ball from a source location (ARG2) to a target location (ARG5) on the field in the intention of putting a team-mate (ARG3) in a position to shoot at goal. Typically, the source location of a cross is somewhere near the byline, and its target location is somewhere near the opponent's goal.

Examples:
(1) [Ronaldo]ARG1 delivered a **cross** [from the by-line]ARG2 [for Milan Baros]ARG3
(2) [Jørgensen]ARG1 put over a **cross** [with the outside of his right foot]ARG4 [for Jon Dahl Tomasson]ARG3

(3) [Quaresma]ARG1 swung an inviting **cross** [into the box]ARG5 which was deflected on to Maniche.

Figure 1: Lexical description of the LU 'cross"

dispossess.v

In a one-on-one challenge at a certain location on the field (ARG3), the attacking player (ARG1) manages to take the ball from the player in possession (ARG2).

Examples:

(1) [Benayoun]ARG1 was tripped after **dispossessing** [Costas Kaiafas]ARG2 [on the edge of his own area]ARG3.

(2) On 16 minutes Hungary went close when [Robert Waltner]ARG2 **was dispossessed** [by Maltese goalkeeper Justin Haber]ARG1 at the last gasp.

(3) [Ronaldo]arg1 **dispossessed** [Wisla goalkeeper Radoslaw Majdan]ARG2 [on the edge of the box]ARG3 only for Arkadiusz Glowacki to produce a last-ditch tackle.

(4) PSV's energy and endeavour was enthralling, with [Park]ARG1 typifying their approach by **dispossessing** [Andrea Pirlo]ARG2 [on the centre spot]ARG3 in the 28th minute and releasing countryman Lee Young-Pyo on the left.

Figure 2: Lexical description of the LU 'dispossess'

tunneln.v

In a one-on-one challenge at a certain location on the field (ARG3), the player in possession (ARG1) manages to overcome the attacking player (ARG2) by playing the ball between the latter's legs.

Examples:

(1) [Diogo Rincón]ARG1 **tunnelte** [Paul Freier]ARG2 [im Strafraum]ARG3 und sein Schuss trudelte an Jörg Butt vorbei und landete in Netz.

(2) [Ailton]ARG1 **tunnelte** [Chris]ARG2 [an der Strafraumgrenze]ARG3 und spielte so Klasnic frei.

(3) [Auf der linken Seite]ARG3 geht [der Angreifer]ARG1 auf und davon, **tunnelt** [Lucio]ARG2 und schnibbelt das Leder gekonnt ins rechte untere Eck (44.).

(4) In der 10. Minute **tunnelte** [Arvidsson]ARG1 [den Ex-Bochumer Fahrenhorst]ARG2, verzog aber aus kurzer Distanz.

Figure 3: Lexical description of the LU 'tunneln'

As the following table illustrates, so far (March 2006) more than 1,200 lexical units have been described in this way[2]:

	DE	EN	FR	Total
LUs	554	383	286	1223
Nouns	277	172	142	591
Verbs	263	196	135	594
Examples	2292	1627	1300	5220

Table 1: Lexical units and examples in the soccer frame net

The part of the vocabulary that has been most extensively analyzed so far are words describing individual events during a match (shots, passes, goals etc.). Whereas the resource seems to be relatively complete in this area in so far as the corpus only infrequently uncovers LUs that have not yet been accounted for, other areas of the vocabulary have not yet been analyzed with the same amount of detail. Most importantly, this regards words that speak about a match as a whole (and its place in a competition) and words that denote actors and objects of a match (e.g. goalpost, penalty area, etc.). It is expected that a complete analysis of these areas of vocabulary will at least double the existing number of LUs.

4. Ontologies for lexicographic purposes

Prévot et al. (2005) distinguish three different options for linking ontologies and lexical resources: (1) *restructuring* a computational lexicon on the basis of ontological-driven principles; (2) *populating* an ontology with lexical information and (3) *aligning* an ontology and a lexical resource. In the SFN, the first of these options is explored – my interest in ontologies is mainly concerned with their ability to provide additional layers of structure to a dictionary. From the dictionary user's point of view, these additional layers of structure should provide a means of navigating the vocabulary that goes beyond traditional lexicographic access structures (the two most important of which are alphabetical lists of head words and thesaurus-like groupings of sense related words).

The most straightforward way of linking lexicographic data to an ontology for lexicographic purposes is to assign individual lexical units to specific members of a well-defined system of (possible interrelated) language-neutral concepts. In this way, various types of semantic equivalence between two different lexical units can be expressed.

1) grouping synonymous words: the fact that two lexical units are synonymous can be expressed by assigning them to the same concept in the ontology. For instance, the English nouns "penalty" and "spot-kick" will be mapped to one and the same concept PENALTY_KICK in the ontology.

2) grouping semantically equivalent predicates of different part-of-speech types: the same principle can be applied

2 The fact that German LUs are significantly more numerous than English and French LUs is partly due to the different corpus sizes (see above), but partly also to the tendency of German to form complex compounds that enter as individual LUs into the resource.

also across different part-of-speech categories. For instance, the noun "through-ball" and the verb "to release" both carry the core meaning of "(playing) a long pass such that its recipient can get through on goal". Linking both these lexical units to a concept THROUGH-BALL in the ontology captures this.

3) distinguishing polysemous words: conversely, the polysemy of a given lemma can be captured by assigning the different uses to different concepts in the ontology. Thus, for instance, one use of the French verb "marquer" would be assigned to a concept MARK_PLAYER, while another use would be assigned to a concept SCORE_GOAL.

4) cross-linguistic linking: just like an ontology can be used to capture synonymy within a language, it can also be used as an interlingua for representing translation equivalence across languages. For instance, the fact that the English lexical unit "hat-trick" translates as "Hattrick" into German and as "coup du chapeau" into French can be represented by assigning all three units to a concept HAT_TRICK in the ontology.

Clearly, this way of interfacing lexical data with an ontology covers a substantial part of the information one would expect of a traditional mono- or bilingual dictionary. The ontology, in this case, is simply a language-neutral meta-structure that is used to indirectly capture those relationships that a traditional dictionary would express by direct links between synonymous or otherwise semantically equivalent lexical units. As Alexa et al. (2002) point out, such a manner of proceeding can have great practical value in dictionary creation and maintenance. For instance, with a language-neutral ontology as a backbone to one or several monolingual lexicographic resources, it may become easier for a lexicographer to construct the same resource for an additional language. From the user's point of view, however, these types of links alone do not yet constitute a substantially novel way of working with a dictionary. The next three sections will illustrate ways of interfacing lexical resources with ontologies that may be more innovative in that respect.

5. Poly-hierarchy of concepts

Mapping the lexicographic descriptions exemplified in section 1.3. to a set of ontology concepts as described in section 2 results in a list-like organization of the lexicon as in table 2.

Additional structure is established by organizing concepts into a poly-hierarchy, i.e. by adding links between them that are to be interpreted as an "is_a" relation. For the set of concepts in this example, the most obvious such link is that between SET-PIECE as a superordinate concept of all other concepts - a SET-PIECE is, by definition, the general term for bringing the ball back into play after some kind of interruption. Depending on the type of interruption, this will be a CORNER, a FREE-KICK, a PENALTY etc. Introducing the types of interruption as intermediate concepts yields the hierarchy depicted in figure 4.

Concept	***EN***	***DE***	***FR***
CORNER	*corner*	*Eckball, Ecke, Eckstoß*	*corner c. d. p. de coin*
FREE-KICK	*free-kick*	*Freistoß*	*coup franc*
GOAL-KICK	*goal-kick*	*Abstoß*	*c. d. p. de but*
PENALTY	*penalty spot-kick*	*Elfmeter, Elfer Strafstoß*	*penalty, c. d. p. de réparation*
PUNT	*punt (n), punt(v)*	*Abschlag, abschlagen*	*dégagament*
SET-PIECE	*set-piece dead ball position*	*Standard, Standardsituation ruhender Ball*	*coup de pied arrêté*
THROW_OUT	*throw out*	*Abwurf, abwerfen*	*renvoi de la main*
THROW-IN	*throw-in, throw*	*Einwurf, einwerfen*	*touche*

Table 2: A list of concepts with corresponding lexical units

SET-PIECE
- AFTER FOUL
  - **PENALTY**
  - **FREE-KICK**
- AFTER BALL OFF FIELD
  - AFTER BALL OVER GOAL LINE
    - **CORNER**
    - **GOAL-KICK**
  - AFTER BALL OVER TOUCH LINE
    - **THROW-IN**
- AFTER GOALKEEPER CONTROLS BALL
  - **PUNT**
  - **THROW-OUT**

Figure 4: The SET-PIECE concept and subordinated concepts

With this kind of hierarchy, the dictionary user is given a means of discovering semantically closely related lexical units. For instance, by navigating the hierarchy, he is able to learn that "dead ball position" is a hyperonym of "corner" and that "corner" and "throw-in" are co-hyponyms.

However, this is not the only possible way of organizing the given concepts. Other useful distinctions are:

1) Set-pieces that are carried out by shooting the ball (CORNER, FREE-KICK, GOAL-KICK, PENALTY, PUNT) vs. set-pieces that are carried out by throwing the ball (THROW_OUT, THROW-IN)

2) Set-pieces that are awarded by the referee (CORNER, FREE-KICK, GOAL-KICK, PENALTY, THROW-IN) vs. set-pieces that are not (THROW_OUT, PUNT)

3) Set-pieces that can be conceived as a pass (i.e. that may have a team-mate as a potential recipient: CORNER, FREE-KICK, GOAL-KICK, PUNT, THROW_OUT, THROW-IN) and set-pieces that can be conceived as a shot (i.e. that can be directed directly at goal: PENALTY and, again, FREE-KICK)

Representing these distinctions as additional concepts and adding hierarchical links accordingly should also be helpful to the dictionary user to understand semantic differences and commonalities between the LUs directly associated with the superordinate term "set-piece".

6. Semantically typing arguments

The basic lexicographic building block of the SFN does not only consider the lexical unit itself, but also its arguments (see section 3.4). Consequently, the linking of the lexical resource to an ontology can also be done for the arguments of a predicate. For instance, the three arguments of the LU "to flick on" can be assigned to the concepts PASS and PLAYER in the ontology, as in the following annotated example:

(4) [A diagonal ball from Ioannis Christou]PASS **was flicked on** [by Thomas Makris]PLAYER [to Chloros]PLAYER

Likewise, the arguments of the LU "to award" are assigned to the concepts TEAM, COMPENSATION and OFFENSE in the following annotated example:

(5) On 71 minutes [Terek]TEAM **were awarded** [a penalty]COMPENSATION [after Mariusz Mowlik's handball]OFFENCE, but Khomukha's spot-kick was weak and Piatek easily parried.

As a rule, the concepts suitable to be assigned to an argument of LUs will be more general than the concepts assigned to the LUs themselves, i.e. they will usually be nodes that are relatively high up in the concept hierarchy. In fact, it has been found that the large majority of all arguments can be covered by no more than 25 different concepts, the most common of which are concepts such as PLAYER, BALL, LOCATION, PART_OF_BODY etc.

In terms of dictionary use, these types of links from the lexicographic resource into the ontology offer an important new way of navigating the vocabulary. Consider again the LU "free-kick" which has been assigned to the concept FREE-KICK as described in section 3. The poly-hierarchy of concepts will provide the information that a) a free-kick can be conceived as a kind of PASS and that b) a free-kick is a COMPENSATION awarded by the referee after a foul. Since the arguments of the LUs "to flick on" and "to award" are assigned to the same concepts, a user looking up "free-kick" thus has a simple means of discovering not only the meaning of the term itself, but also of learning about other predicates with which it is used as an argument.

Following the same principle, a lookup of an LU like "goalkeeper" will not only reveal that this is a word used to describe one of the actors of a soccer team, that it is synonymous to the LU "keeper", that "player" is one of its hyperonyms and "defender", "playmaker" etc. its co-hyponyms and that it translates as "Torhüter" into German and "gardien" into French, but also that "to punt", "to punch", "to spill", "to fist", "parade" and "save" are LUs that take an argument of the type GOALKEEPER.

7. Scenarios

The poly-hierarchy of concepts described so far is exclusively concerned with static semantic relations between lexical units. However, a soccer match being a dynamic event unfolding over time, temporal relationships between concepts also play an important role for organizing soccer vocabulary. To describe such temporal relationships, the FN methodology offers the concept of a scenario, i.e. a background description for a sequence of events and transitions. Reidsma et al. (2003), in their ontology-based approach to multimedia information extraction from soccer data, use a similar notion which they call "scene".

In the SFN, a number of prototypical sequences of events in a soccer match have been described as scenarios. These scenarios are all centered around a core event (e.g. a shot) which has a number of participants, and which may be composed of smaller substages. In addition to that, a scenario describes background prerequisites that are necessary for the core event to happen, as well as possible outcomes or following actions. As an example, consider the description of the pass scenario. The main participants in a pass are the passer, the recipient, the ball, a source and a target location on the field, as well as a potentially intervening player (the interceptor) and a potential second recipient. The following diagram illustrates how they take part in a passing event.

The core event of this scenario is lexicalized by LUs such as "to pass", "to center", "through-ball" and "cross", and the arguments of these LUs are linked to the corresponding participants of the scene:

(6) With three minutes remaining [substitute Marcelo Zalayeta]PASSER passed [the ball]BALL [into the middle]TARGET where the unmarked Trezeguet made it 4-1.

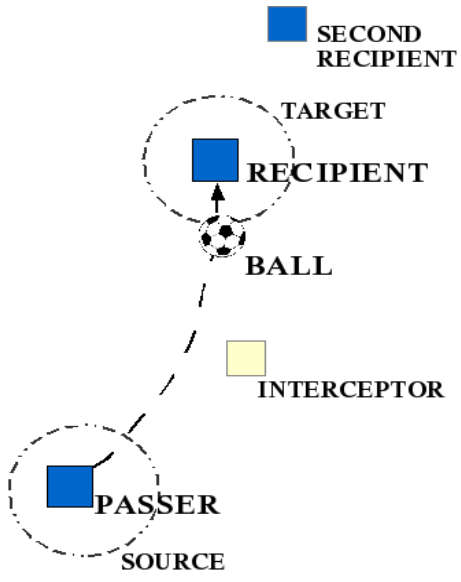

Figure 5: The pass scenario

Note that this assignment is different from the semantic typing of arguments described in the previous section, the difference being basically one between types and roles: whereas the argument "substitute Marcelo Zalayeta" in sentence (6) would be assigned the semantic type PLAYER, a property which holds independently of a specific scene, the assignment of the role PASSER is only valid within this particular passing event.
The same assignment is applied to LUs that do not describe the core event, but a substage or an outcome of it. For instance, one possible outcome of a pass event is that the recipient controls a pass. This is lexicalized by LUs such as "to chest down", "to control", "to fasten on" etc. Linking the arguments of these LUs to the concepts describing the participants in the pass scenario yields annotations of the following type:

(7) [Jeff Whitley]RECIPIENT chested down [a free-kick from Mark Clyde]PASS [at the edge of the box]TARGET
(8) [He]RECIPIENT fastened on [to Shearer's lay-off]PASS [20 metres out]TARGET

Besides controlling a pass, other substages or possible outcomes in the pass scenario include connecting with a pass (LUs: to connect, to meet etc.), missing a pass (LUs: to miss, to miscontrol etc.) flicking on a pass (LUs: to flick on, flick-on) and intercepting a pass (LUs: to intercept, interception etc.). The descriptions of these LUs are linked to the pass scenario in the same way:

(9) [García]RECIPIENT flicked on [Steven Gerrard's set-piece]PASS [for centre-back Hyypiä]SECOND_RECIPIENT
(10) Then [Gert Verheyen]INTERCEPTOR intercepted [a Shakhtar pass]PASS and fed Balaban.

Systematically applying this kind of link between the lexical data and a language neutral description of scenarios provides one further way for the dictionary user to discover semantic relations between lexical items. For instance, starting with a look-up of the LU "pass", the network of links belonging to the pass scenario will take the user to other lexical units describing events that are temporally related to this LU. Since talking about soccer prototypically means lexicalizing sequences of events, this should be of great practical value especially when a dictionary is used actively, i.e. to produce rather than merely to understand a linguistic expression.
Moreover, if the dictionary's task is to help the user to *translate* from one language into another, this kind of information can be crucial in dealing with lexical gaps. Consider, for instance, the following sentence which contains two LUs that would be assigned to different parts of the pass scenario – the verb "to connect" and the noun "cross":

(11) [Bresciano]RECIPIENT missed the target after **connecting** [with [Fabio Simplicio's]PASSER **cross** [from the left]SOURCE]PASS

German does not offer a straightforward translation equivalent for the verb "to connect". However, knowing that the subject of this verb describes the participant RECIPIENT of the pass scenario allows the user to reformulate the sentence by integrating the recipient role as an argument of the LU "Flanke" (which, in turn, is marked as a translation equivalent of "cross" via the concept mapping described in section 2). In that way a translation like the following one might be derived:

(12) [Fabio Simplicio]PASSER schlug eine **Flanke** [von links]SOURCE [auf Bresciano]RECIPIENT. Dieser verfehlte jedoch das Ziel.

8. Summary and Outlook

This paper has presented on-going work on a multilingual lexical resource of soccer language based on frame semantic principles. It has been sketched how different links from the description of lexical units and their arguments into different systems (a poly-hierarchy and a set of scenarios) of language-neutral concepts can act as an access structure to the resulting dictionary, and it has been argued that this kind of access structure provides the user with novel ways of discovering and exploiting semantic relationships between words that traditional dictionaries do usually not cover.
The work as presented here is far from being complete. The next step in the development of the SFN will therefore be to increase the number of lexical units and to supplement the concept hierarchies and scenario descriptions accordingly. Following that, a very important objective will be to develop user interfaces that allow the dictionary user to actually exploit in practice the type of links between lexicographic and ontological data described here.
Concerning the lexicographic side of the work, a more long-term goal is to supplement the corpus data, which at the moment consists entirely of written match reports, with *spoken* data. It is expected that this will not only lead to a substantial number of new lexical units (because spoken soccer language, even more than its written counterpart, is known to be very rich in idiomatic expressions), but also that it will reveal new argument patterns for existing LUs. Furthermore, adding audio data to the lexicographic description of LUs has an obvious didactic value especially for a foreign language user of a dictionary. A number of audio recordings of German radio soccer commentaries have been collected as a first step towards this goal.
Concerning research into ontologies, no concrete steps beyond the ones sketched here are planned for the near future. However, there are some obvious ways in which this work could be related to other studies whose focus is more on natural language processing than on lexicography for human users: Firstly, just like ontologies are constructed here to structure a given set of lexical units, these lexical units could conversely be used to populate existing ontologies with lexical material. This would correspond to the second type of interface between lexical resources and ontologies described by Prévot et al. (2005).

It could be interesting to investigate how well the bottom-up method of collecting LUs and then using an ontology to structure them fits with a top-down method of devising an ontology for a given domain and then "filling" it with lanuage-specific information. Secondly, a more formalized approach to ontology modelling than the one presented here might be a future line of research. The ontologies in the SFN are formulated as simple XML files with pointers into the lexical data, containing no more information than what is directly needed for the lexicographic task at hand. Expressing the same ontologies in a standardized framework, adding rules about concepts and linking concepts to upper ontologies like SUMO could constitute a way of making the knowledge contained in the SFN usable for machine processing purposes.

9. Acknowledgments

The work presented in this paper is being funded by a post-doc grant from the German Academic Exchange Service (DAAD). I am grateful to the Berkeley FrameNet project (Charles Fillmore, Collin Baker, Michael Ellsworth, Josef Ruppenhofer, Miriam Petruck) and its current visitors (Kyoko Ohara, Carlos Subirats, Jan Scheffczyk) for their support.

10. References

Melina Alexa, Bernd Kreissig, Matrina Liepert, Klaus Reichenberger, Lothar Rostek, Karin Rautmann, Werner Scholze-Stubenrecht, and Sabine Stoye. 2002. The Duden Ontology: An Integrated Representation of Lexical and Ontological Information. In *Proceedings of The Ontologies and Lexical Knowledge Bases Workshop (OntoLex02), Las Palmas*.

Hans C. Boas. 2006. Semantic Frames as Interlingual Representations for Multilingual Lexical Databases. In *International Journal of Lexicography 18(4), 445-478.*

Paul Buitelaar, Michael Sintek, and Malte Kiesel. 2005. Feature Representation for Cross-Lingual, Cross-Media Semantic Web Applications. In *Proceedings of the Workshop on Knowledge Markup and Semantic Annotation (SemAnnot2005) at the International Semantic Web Conference, Galway, Ireland, Nov. 2005.*

Paul Buitelaar, Thomas Eigner, Greg Gulrajani, Alexander Schutz, Melanie Siegel, Nicolas Weber, Philipp Cimiano, Günter Ladwig, Matthias Mantel, Honggang Zhu. 2006. Generating and Visualizing a Soccer Knowledge Base. In Frank Keller, Gabor Proszeky (eds.): *Proceedings of the EACL06 Demo Session, 4/2006*

Charles Fillmore. 1982. Frame semantics. In *Linguistics in the Morning Calm*, Seoul, Hanshin Publishing Co., 111-137.

A. Nijholt, R. op den Akker, and F. de Jong. 2003. Language Interpretation and Generation for Football Commentary. University of Twente.

Laurent Prévot, Stefano Borgo, and Alessandro Oltramari. 2005. Interfacing Ontologies and Lexical Resources. In *Proceedings of OntoLex2005, Jeju Island, South Korea.*

Dennis Reidsma, Jan Kuper, Thierry Declerck, Horaccio Saggion, and Hamish Cunningham. 2003. Cross document ontology based information extraction for multimedia retrieval. In *Supplementary proceedings of the ICCS03, July 012003, Dresden, 2003.*

Josef Ruppenhofer, Michael Ellsworth, Miriam Petruck, and Chris Johnson. 2005. FrameNet: Theory and Practice. http://framenet.icsi.berkeley.edu/book/book.html

Horacio Saggion, Hamish Cunningham, Kalina Boncheva, Diana Maynard, Cris Ursu, Oana Hamza, and Yorick Wilks. 2002 Access to Multimedia Information through Multisource and Multilanguage Information Extraction. In B. Andersson, M. Bergholtz, P. Johannesson (eds.): *Natural Language Processing and Information Systems: 6th International Conference on Applications of Natural Language to Information Systems, NLDB 2002, Stockholm, Sweden, June 27-28, 2002*.

Dieter Seelbach. 2001. Das kleine multilinguale Fußball-Lexikon. In Bisang, W. & Schmidt, G. (ed.): *Philologica et Linguistica. Historia, Pluralitas, Universitas. Trier.*

Dieter Seelbach. 2002. La traduction des verbes avec adverbes appropriés et des verbes à particule allemands. In *Traduire au XXIème siècle: Tendances de perspectives, Facultés des lettres UATH Athens.*

Dieter Seelbach. 2003. Separable Partikelverben und Verben mit typischen Adverbialen. Systematische Kontraste Deutsch-Französisch / Französisch-Deutsch. In Seewald-Heeg, Uta (ed..): *Sprachtechnologie für die multilinguale Kommunikation. Beiträge der GLDV-Frühjahrstagung.*